DOCUMENT RESUME

ED 157 282 EC 111 767

AUTHOR Alzate, Gloria
TITLE Analysis of Testing Problems in Spanish-Speaking Children.
PUB DATE Jun 78
NOTE 12p.; Paper presented at the World Congress on Future Special Education (First, Stirling, Scotland, June 25 - July 1, 1978)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
DESCRIPTORS Culture Free Tests; Disadvantaged Youth; Elementary Secondary Education; Minority Groups; *Spanish Speaking; Student Evaluation; *Test Bias; *Testing Problems

ABSTRACT

The paper reviews the literature on the problem of testing Spanish-speaking children with North American Instruments, and suggests some solutions to the problem. In the first section, studies done with Spanish-speaking children are considered; studies of other cultures follow; and a final section probes the issue of test bias. (Author/SBH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

173 #1

GLORIA Alzate
Psychologist & Master Sp. Ed.
Calle 64 # 7-37 Apto 801
Bogotá 2, colombia
South America 1

ED157282

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Analysis of Testing Problems in Spanish-Speaking Children

One of the biggest problems of education in Colombia and other Latin American countries is the scarcity of educational and psychological instruments to assess students in intelligence, aptitudes, attitudes, and personality. The lack of resources and time to develop instruments has resulted in translated versions of United States tests and their use with norms and scoring procedures appropriate for the sample on which the tests were standarized. These translations have been used without considering cultural differences that require a complete re-norming, before an instrument can be used to evaluate Colombian students.

Although there are some cultural similarities between countries, there are a lot of differences that do not permit the use of identical indices for group and individual comparisons (Bertou and Clasen, 1971). Culture makes children to think, react, value, believe and act in certain modes, and learning may be conditioned by cultural differences (Jaramillo, 1972). These differences must be taken into consideration in the application of tests, because they will influence children's performance.

The objective of this paper is to review literature on testing problems of Spanish-speaking children with North American instruments, and to suggest some solutions to the problem. First, I will review studies done with Spanish-speaking children and then, studies of other cultures followed by a consideration of test bias.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C.E.C. ©

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

EC111767

Studies with Spanish-Speaking Children

Some studies have evaluated Spanish versions of United States test.

The Catell Sixteen Personality Factor Test was translated into Spanish and was administered to 524 freshmen at Venezuela's Central University (Berton and Clasen, 1971). The researchers found that only four factors had acceptable reliability. They changed the weighted values of the responses, but they could not improve the statistical indices. They concluded that the translated version was not reliable and that cultural differences influenced the performance; cognitive characteristics which influence item responses, and the translation of individual item, should be taken into account in order to adapt the test.

The forms A and B of the PPVT were translated into Spanish and were applied to 33 students of 1 - 6 grades randomly selected of school population of Xalapa (Mexico) (Simon & Joiner, 1976). A new version was constructed by selecting the items in the two forms with high correlation with the total score, high correlation with cronological age, or pass/fail ratio close to .50. The new test was applied to 120 students. The first version was less reliable than the original; the second was better than the first one, and the validity and reliability were improved by item-selection and reordering. The authors concluded that the reliability of the test is affected by translations, and it may be more affected in translations of languages less related than English and Spanish (Simon & Joiner, 1976).

Anastasi and DeJesus (1953) studied the I.Q. of Puerto Rican children living in New York City. Their performance in the Goodenough-Harris Drawing Test was not significantly different from the mean established for whites and blacks of the same neighborhood. Other studies in Spanish or English have shown low

I.Q. of Puerto Rican children. Anastasi and DeJesus concluded that the language is an important factor in I.Q. tests, because the children performed better in the non-language test.

Brelan (1974) compared the performance on six tests of these groups: Afro-American, Mexican-American, Puerto Rican, other Latin American, and whites living in different regions of the United States. He found cross-cultural instability in the mental test items, and questions with difficult verbalization were very difficult for Latin American people.

Meeker and Meeker (1973) analyzed the implications of inadequate testing procedures, such as, deficiencies in intelligence testing, cultural bias, construct validity, and diagnostic utility. They compared Mexican-American and Black, with Anglo groups and they found that intelligence tests are not adequate to assess non-Anglo and lower socio-economic groups. The test procedures are "penalizing" these groups.

Matuszek and Thomas (1972) studied achievement in several reading readiness test of six groups of children: Black, Mexican-American, and Anglo-American, in low and middle class. They found that the test differed markedly across the groups and the subtests did not seem to measure the same factors. They concluded that these tests could not be used interchangeable with children of different socio-economic status and racial-ethnic groups.

The results of all these studies suggest that a United States test cannot be used to assess Spanish-speaking children, unless an adaptation of the test is produced and new norms established. The general agreement is that standarized intelligence tests and achievement tests present linguistic, cultural, and psychological difficulties for Spanish-speaking children in terms of internal and intervening variables, such as, the language of the administrator, the extent

of language factor in them, and cultural differences. A verbal factor militates against the optimal performance of these children. Tests most dependent on English language skills generally resulted in poorest performance, indicating a language problem rather than a learning disability (Zirkel, 1972).

Studies with Different Cultures

There are some studies in different languages and cultures that are interesting and valuable to consider. They have revealed the same results as investigations with Spanish-speaking children and show the importance of language pattern and cultural equivalences in adaptations of instruments.

Bushan (1974) translated the Otis-Quick-Score Mental Ability Test into French. He applied the test to a sample, and then, he submitted the items and distractors to statistical analysis. A review version was applied to a second sample, and he found that the validity of the second version was greater than in the first one. The reliability was almost the same. He concluded that a good test cannot be constructed simply by translating from one language to another, unless studies of validity are conducted. The test should be applied, modified, and standarized for the new culture. This result is a replication of Simon's and Joiner's study with the PPVT, using a translated version from English to Spanish.

Tierney (1974) studied the generalization to other cultures of psycholinguistic findings in syntax and reading comprehension. He worked with a group of Australian students in reading comprehension of materials written at different levels of syntatic maturity. He found that it is necessary to be very careful in generalizing American research findings to another culture, especially in language development of children.

Clark (1965) used a Portuguese translation of one Spanish test, and he applied the test to a group of peace corps trainees. He studied the statistical equivalence of the test and he assumed that there was equivalence when the items showed equal reliability, difficulty, and item-total test correlations. He found statistical consistencies and the translation/adaptation was successful because the common linguistic pattern of the two languages. He recommended that researchers be careful using translations with less related languages.

These investigations show the necessity of studying language patterns and linguistic equivalence in the adaptation of a test from one culture to another. When the languages are very similar the translation will not affect the results, but when the languages differ we must study these differences and introduce changes in the structure of the sentences in the test. Cultural differences should be considered in the content of the test in order to improve the instrument's validity.

Test Bias

A number of studies have been done in the United States in the area of test bias. These investigations have drawn important conclusions that should be considered in the development of instruments for Latin American countries.

Hoepfner and Strickland (1972) studied the problem of test bias in different socio-economic and racial-ethnic groups. Bias was defined as "an item by race interaction in an analysis-of-variance design." They applied the paragraph meaning subtest of the Stanford Achievement Test to a group of whites, blacks, Mexican-Americans, and Orientals. Item regression was obtained but test bias was not conclusively demonstrated in a wholly satisfactory manner,

because socio-economic status, intelligence, and other variables were not controlled. They conclude that when a test designed for one group is given to another it is always possible test bias to exist. Content validity will be different for different groups. The mean for racial minority groups is lower and the standard deviation smaller than those for majority white groups.

Green (1971) conducted an item analysis of the California Achievement Tests using several subgroups of the standarization sample. The sample included urban, and suburban groups of different ethnic groups. The best half of the items in each test were selected for each group. Thirty percent of the items in the upper half of the distribution of the item-test correlations for a group did not meet this criterion with another group. The three suburban groups (minority groups) were relatively similar and the scores were better on the selected items. Green concluded that standarized tests are adequate for majority groups and are biased against minority groups.

Pascale and Jakubovic (1971) reviewed the formats and psychometric rationale of culture-free tests, such as, Raven's Progressive Matrices, Lorge-Thorndike, Catell's Culture Fair Test, Goodenough-Harris Drawing Test, and Davis-Elles games. The authors had various arguments against culture-free tests: 1) The fluctation of I.Q. scores from one age to another shows that I.Q. is not a reliable concept, because it can change depending upon grade level at which it is measured and normed, 2) The test does not provide a valid measure of individual performance, it provides a testing valid for groups, and 3) Test should be analyzed according with cultural grouping's overall performance.

Rogers (1970) in his book, Fallacies of Testing, analyzes not only test

bias, but also testing problems, especially in reading tests. He argues that standarized tests are not appropriate to assess individual achievement, because they compare the performance of the individual with the group and they are usually anchored in the middle of the ability range. Criterion-referenced tests should be used because they are based in the specific skills necessary to achieve a goal and they can diagnose individual problems in order to plan instructional programs. He gives four solutions for assessment: 1) Criterion-referenced tests, 2) Average levels of achievement of some age group as standards of achievement, 3) Norm referenced test with specific behavioral objectives, and 4) Guidelines to develop situation specific criterion measures. Each situation would need its own criterion-referenced test, depending upon socio-economic, geographical and cultural factors.

These investigations have found three principal ideas: 1) Standarized tests are adequate to assess majority groups, but they produce low scores in minority groups, 2) Norms are not adequate to evaluate the performance of groups different from the population on which the test was standarized, and 3) Before using an instrument to assess student abilities, we will have to know if the purpose is to compare the individual with the group or to determine intraindividual differences. In the former case, standarized test can be used, in the later, criterion-referenced tests are better.

Conclusions

Data from different investigations show that translated versions of a North American test should not be used with Latin American children, because the test is less reliable and valid in this population. Adaptations of a test are more appropriate but a number of factors should be considered before it can be used. One of the principal factors is language and linguistic patterns:

the equivalence of syntatical and grammatical factors have to be studied. Other aspects are cultural differences and cognitive characteristics that will affect not only the content validity of the test, but also the perception and the feelings of the children in the exam situation. The test should be applied in order to conduct analysis of items, reliability and validity analysis, and finally to establish new norms for the population in which we plan to use the test.

A representative sample should be selected, including majority and minority groups of rural, suburban, and urban areas, and different socioeconomic levels, because the norms should be established taking into consideration all the different groups of the population.

Due to lack of resources, Latin American countries are unable to conduct the necessary investigations for developing adaptations or standarizations of tests. Perhaps one solution to the problem is to start with criterion-referenced tests that are easier to develop and provide accurate assessment of individual abilities, and in the future, to conduct standarization studies if the interest is to make an interindividual comparison. North American tests can be used as a model in the identification of the sequence of skills and the goals of the learning process. Another solution is to begin developing the most necessary standarized tests, such as, intelligence and ability instruments, and then to produce tests in other areas.

This paper has tried to clarify some testing problems in Spanish-speaking children but furthermore investigation is needed in order to identify cognitive, linguistic, and other cultural characteristics that may affect the children's response to a test situation. These factors should be studied not only

between different countries, but also within each country, because there are regional characteristics that may produce ideological and behavioral differences.

References

Anastasi, A. & DeJesus, C. Language development and nonverbal I.Q. of Puerto Rican preschool children in New York City. Journal of Abnormal and Social Psychology, 1953, 48, 357-365.

Bertou, P. & Clasen, R. An analysis of a Spanish translation of the Sixteen Personality Factors Test. Journal of Experimental Education, 1971, 39, 13-21.

Bhushan, Vidya. Adaptation of an intelligence test from English to French. Journal of Educational Measurement, 1974, 11, 43-48.

Breland, H. H. An Investigation of Cross-Cultural Stability in Mental Tests Item. Chicago: Annual meeting of the American Educational Research Association, 1974. (ERIC Document Reproduction Service No. ED 093 916)

Clark, J. Item by item parallel achievement test for different foreign languages. Journal of Educational Measurement, 1965, 2, 77-83.

Green, D. R. Racial and Ethnic Bias in Test Construction. Monterey, Calif.: Final Report. CTB/McGraw Hill, 1971. (ERIC Document Reproduction Service No. ED 056 090).

Hoepfner, R. & Strickland, G. P. Investigating Test Bias. Los Angeles: California University, 1972. (ERIC Document Reproduction Service No. ED 066 443).

Jaramillo, M. L. Cultural Differences Revealed Through Language. New York: Columbia University, 1972. (ERIC Document Reproduction Service No. ED 066 522).

Matuszek, P. A. & Oakland, T. D. A Factor Analysis of Several Reading Readiness Measures for Different Socioeconomic and Ethnic Group. Chicago: Annual meeting of the American Educational Research Association,

1972 (ERIC Document Reproduction Service No. ED 068 909).

Meeker, M. & Meeker, R. Strategies for Assessing Intellectual Patterns in Black, Anglo, and Mexican-American Boys -- On Any Other Children -- and Implications for Education. Not availeble, 1973 (ERIC Document Re-Production Service No. ED 084 909).

Pascale, P. J. & Jakubovic, S. The Impossible Dream: A Culture-Free Test. Not available, 1971 (ERIC Document Reproduction Service No. ED 054 217).

Renzulli, J. S. & Paulus, D. H. A cross-validation study of the item ordering of the Peabody Picture Vocabulary Test. Journal of Educational Measurement, 1969, 6, 15-20.

Rogers, F. The Fallacies of Testing. Bloomington, Ind.: Conference on reading and the national interest, 1970 (ERIC Document Reproduction Service No. ED 040 020).

Simon, A. J. & Joiner, L. M. An American version of the Peabody Picture Vocabulary Test. Journal of Educational Measurement, 1976, 13, 137-143.

Tierney, R. J. The Generalization of Psycholinguistic Research to Other English Speaking Cultures. New Orleans: Annual meeting of the International Reading Association, 1974. (ERIC Document Reproduction Service No. ED 099 922).

Zirkel, P. A. Spanish-speaking Students and Standarized Tests. Albany: State Education Department, 1972 (ERIC Document Reproduction Service No. ED 080 594).